



Short communication

Methods for comparing multiple digital PCR experiments



Michał Burdukiewicz^{a,*}, Stefan Rödiger^{b,1}, Piotr Sobczyk^c, Mario Menschikowski^d, Peter Schierack^b, Paweł Mackiewicz^a

^a University of Wrocław, Faculty of Biotechnology, Department of Genomics, Wrocław, Poland

^b Institute of Biotechnology, Brandenburg University of Technology Cottbus – Senftenberg, Senftenberg, Germany

^c Wrocław University of Technology, Institute of Mathematics and Computer Science, Poland

^d Dresden University of Technology, Institute of Clinical Chemistry and Laboratory Medicine, Germany

ARTICLE INFO

Article history:

Received 11 December 2015

Accepted 28 June 2016

Keywords:

Digital PCR

dPCR

Generalized Linear Models

GLM

Multiple comparison

ABSTRACT

The estimated mean copy per partition (λ) is the essential information from a digital PCR (dPCR) experiment because λ can be used to calculate the target concentration in a sample. However, little information is available how to statistically compare dPCR runs of multiple runs or reduplicates. The comparison of λ values from several runs is a multiple comparison problem, which can be solved using the binary structure of dPCR data. We propose and evaluate two novel methods based on Generalized Linear Models (GLM) and Multiple Ratio Tests (MRT) for comparison of digital PCR experiments. We enriched our MRT framework with computation of simultaneous confidence intervals suitable for comparing multiple dPCR runs. The evaluation of both statistical methods support that MRT is faster and more robust for dPCR experiments performed in large scale. Our theoretical results were confirmed by the analysis of dPCR measurements of dilution series.

Both methods were implemented in the *dpcR* package (v. 0.2) for the open source R statistical computing environment.

© 2016 The Author(s). Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Digital PCR (dPCR) is a PCR-based method, which enables a precise quantification of nucleic acids. The conventional PCR performs single reaction per one sample. However, in the case of dPCR the sample is separated into a large number of partitions, in which the reaction is carried out individually (clonal amplification). The partitioning allows to assess the number of amplified template molecules by detection of their presence (positive call) or absence (negative call) in particular partitions [1,2]. Since the output of these results is binary, we do not know if the positive partition contains one or more template molecules. The Poisson transformation is required to compute the average number of template molecules per partition, expressed by λ :

$$\lambda = -\log\left(1 - \frac{k}{n}\right) \quad (1)$$

where k is number of positive partitions and n is number of negative partitions.

Thanks to that, it is possible to measure precisely concentrations of nucleic acids with high sensitivity and reliability. Therefore, dPCR found common applications in amplification of DNA samples for next-generation sequencing and detection of variation in genomic sequences, e.g. point mutations and repeats [1].

In contrast to the conventional PCR, in which the number of amplification cycles ideally is proportional to the initial copy number, dPCR does not depend on the cycle number to determine the initial amount of nucleic acids in the sample. In particular, the quantitative real-time PCR is known to be demanding regarding preprocessing, quantification cycle determination and multi-plate measurements [3–6]. The dPCR methodology eliminates the dependence on the exponential shape of data to estimate the concentration of target nucleic acids and enables their absolute quantification. Therefore, this method does not need calibration curves and may even be less susceptible to inhibitors. The amplification chemistry of absolute quantification in the dPCR is orchestrated by well established methods such as analogue PCR or isothermal amplification [7,2,8–10].

Precision, sensitivity, dynamic range, number of partitions and their volume are important parameters in a dPCR system [11]. Moreover, technical replicates are affected by different intrinsic and extrinsic influences increasing the variation of obtained results. This variation needs to be assessed to make a valid statement about

* Corresponding author.

E-mail address: michalburdukiewicz@gmail.com (M. Burdukiewicz).

¹ These authors contributed equally to this work..

the assay performance. As all diagnostic methods, the dPCR requires tools to check consistency of obtained results. There is a growing need for statistical methods for the analysis and design of experiments using digital PCR experiments.

Previously, two methods to compute the λ value and its uncertainty were described. Dube's approach uses confidence intervals [12], whereas Bhat's method is based on the uncertainty [13]. The latter is not a confidence interval in the statistical sense, but nevertheless can be employed to compute probability coverage of the estimated λ value. The Dube's method computes binomial confidence intervals for proportion k/n using the method of normal approximation. Briefly, the binomial distribution of positive counts with the parameters $p = k/n$ and n trials is approximated by a normal distribution. Both Bhat's and Dube's methodologies do not address multiple comparisons of runs, which is a common task during the design and analysis of dPCR experiments.

Here, we propose two approaches for the comparison of multiple dPCR experiments. Both are able to simultaneously compare the λ values of multiple runs. One of them is based on Generalized Linear Models and the second one is the uniformly most powerful ratio test combined with multiple testing correction. Our findings were implemented in the **R** statistical computing environment [14], which has numerous functionalities devoted to analysis of dPCR and qPCR reactions [15].

Methods

Generalized Linear Models – GLM

Generalized Linear Models (GLM) are linear models for data in which the response variable may have a non-normal distribution (e.g. binomial distribution of positive partitions in the case of dPCR experiments). We employ a simplistic model reflecting the relationships between variables in dPCR results, given by formula:

$$\log Y = \beta^T X \quad (2)$$

where Y are counts of positive partitions, X are experiments names (categorical data) and β are coefficients for every run. Since the binomially distributed response is explained by the linear combination of parameters (in our specific case experiment names) we call

such model binomial regression as described in detail elsewhere [16]. Briefly, we employed the logarithm function (function that limits values of response) and the estimated means of copies per partitions by calculating $\hat{\lambda} = \exp \beta$. Importantly, our GLM employs the quasibinomial model, which describes the binomial distribution with excessive zeros. That means that number of zeros may surpass its value predicted by the binomial distribution [17].

The GLM model used in this analysis and represented by Eq. (2) can be refined by adding further effects, such as the technical replication. This may decrease the variance within replicate experiments.

Next the differences between estimated coefficients are evaluated pairwise [18,19]. The resulting p -values require no posterior correction because the familywise error is controlled. This approach is a single-step procedure, because the decision (rejection or acceptance of null hypothesis) is not based on the decision of another hypothesis. As implied by the name, all tests are made in the single step, independently and simultaneously.

Multiple testing

The $\hat{\lambda}$ from two or more dPCR experiments may be pairwise compared using the uniformly most powerful (UMP) ratio test. Uniformly most powerful tests have the highest statistical power (probability that the test correctly rejects the false null hypothesis H_0) for all tests with the same α . The p -values are computed using the TST (twice the smaller tail) method [20,21]. Similarly to the GLM method, this is a single-step approach. However, to control the familywise error rate the pairwise comparison requires an appropriate adjustment of p -values, as the Benjamini–Hochberg correction [22]. The UMP ratio test has the following null-hypothesis:

$$H_0 : \frac{\lambda_1}{\lambda_2} = 1 \quad (3)$$

where λ_1 and λ_2 are mean numbers of template molecules per partition in two experiments. It is also possible to employ other tests (e.g., proportion test) designed to determine the probabilities of having positive partitions [23].

Instead of relying on confidence intervals (CI) computed by the UMP test, we used CIs calculated by the Wilson method. It was

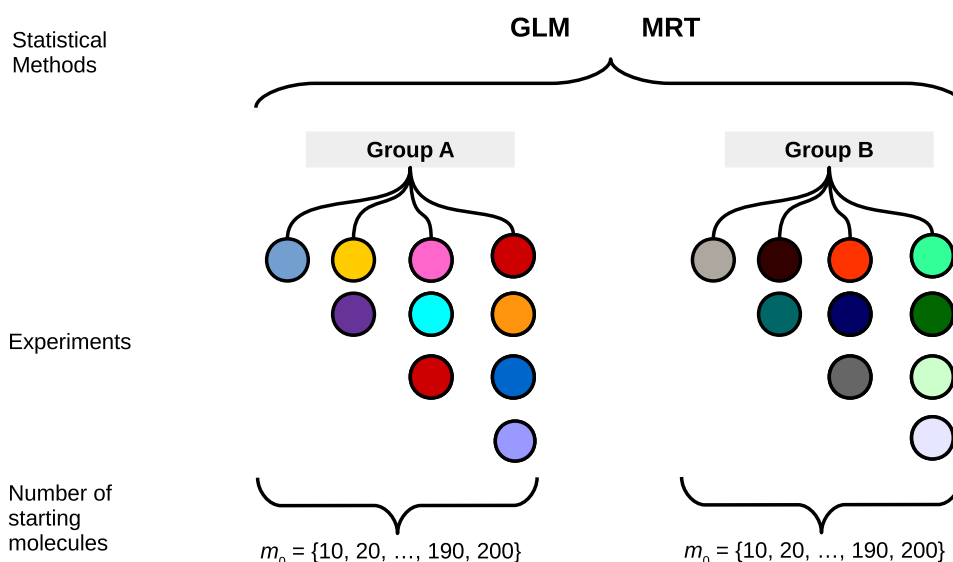


Figure 1. Scheme of the *in silico* dPCR for two approaches based on Generalized Linear Models (GLM) and Multiple Ratio Test (MRT). The groups (A and B) contain one to four runs (coloured circles). Each run within one group is generated using the same number of molecules in 1000 partitions. The number of molecules (m_0) range from 10 to 200. Equally-sized groups were finally compared by GLM and MRT methods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

proven [24] that Wilson's CIs have generally a good coverage probability. In particular, this applies when the probability of success is close to zero, such as in cases of dPCR experiments where the concentration of the template is low. The Wilson's confidence intervals

Table 1

Formulations of uncertainty intervals of compared methods: MRT, Dube and Bhat. Φ : cumulative normal distribution function.

Method name	Formulation
MRT	$-\log \left(\frac{1}{1 + \frac{k}{n} \left(\Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right)^2} \left(\frac{k}{n} + \frac{\left(\Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right)^2}{2n} \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{k^2}{n} \left(1 - \frac{k}{n} \right) + \frac{\left(\Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right)^2}{4n^2}} \right) \right)$
Bhat	$\lambda \pm \sqrt{\frac{\frac{k}{n}}{n \left(1 - \frac{k}{n} \right)}}$
Dube	$-\log \left(1 - \frac{k}{n} \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{\frac{k}{n} \left(1 - \frac{k}{n} \right)}{n}} \right)$

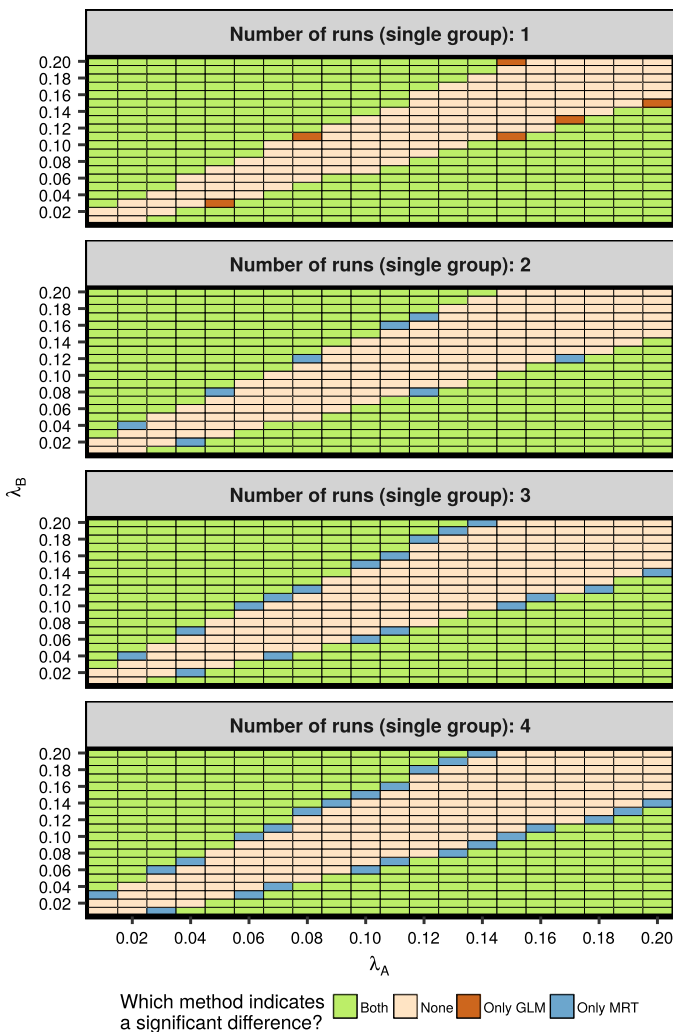


Figure 2. Monte Carlo evaluation of the GLM and MRT methods for number of runs in a single group from one to four. The colour of the tile describes which of these methods distinguished between two compared groups (at the level $\alpha = 0.05$). Abscissa and ordinate represent the average number of template molecules per partition used in the *in silico* experiment in equinumerous groups A and B, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

are calculated independently for every dPCR run and its significance level is adjusted using the Dunn–Šidák correction [25]:

$$\alpha_{\text{adj}} = 1 - (1 - \alpha)^{1/T} \quad (4)$$

where α is the significance level and T is the number of tested hypotheses.

Such intervals are wider than non-corrected CIs but they simultaneously contain the mean number of template molecules per partition of several experiments at a given significance level. Due to this, simultaneous confidence intervals allow testing the results against thresholds while keeping the familywise error at specified level. In other words, the value of λ is statistically different (on the significance level equal to the confidence level) from the threshold value only if the threshold value is not covered by the confidence intervals.

For clarity, we decided to call our novel approach Multiple Ratio Test (MRT).

Confidence intervals

The confidence intervals computed by MRT were compared to uncertainty intervals calculated using two other methods: Bhat's

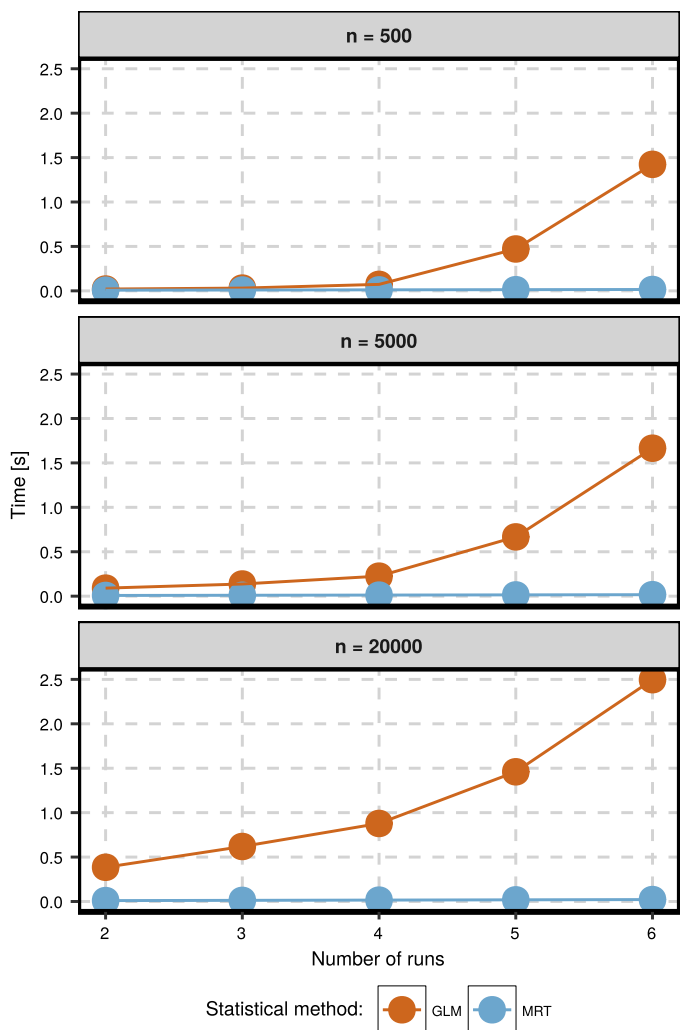


Figure 3. The impact of number of runs on the average time of data analysis by GLM and MRT methods. The MRT outperforms GLM for larger data sets. The performance does not depend on the total number of partitions per run, n . The average time was calculated from 100 repetitions.

[13] and Dube's [12]. The formulation of considered methods is presented in Table 1.

Evaluation of statistical methods

Two implemented approaches, GLM and MRT were compared over 2 millions *in silico* dPCR runs. Each simulation contained two groups with the same number of runs (from one to four runs), which started from the same number of template molecules. The simulation was repeated 60 times (Fig. 1). Both frameworks were used to compare the runs. The GLM approach used a quasibinomial model.

Experiment setup

To validate the proposed MRT analysis framework, we performed a dPCR experiment on dilution series using the `dPCRmethy1` dataset from the `dPCR` package [26]. This dataset contains a dilution of methylated human gDNA (see package for details). For each concentration level, three samples were measured.

Implementation

All findings were implemented in the `dPCR` package (v. 0.2) [26] for the **R** statistical computing environment. The GLM uses functionality from `multcomp` package [27] and the ratio for MRT comes from `rateratio.test` [28]

The stable version of the package is available from <https://cran.r-project.org/web/packages/dPCR/index.html>. Plots were created with `ggplot2` [29]. All analyzes were done in dedicated **R** environments as described previously [15]. An executable **R** script is attached as a supplement.

Results and discussion

Monte Carlo evaluation

The statistical tests showed that neither GLM nor MRT rejected the true H_0 (Eq. (3)) in all cases. Both methods had 'blind spots', in which they could not distinguish runs containing different numbers of template molecules (Fig. 2). The number of non-rejected hypotheses widened with the growth of the λ values. For example,

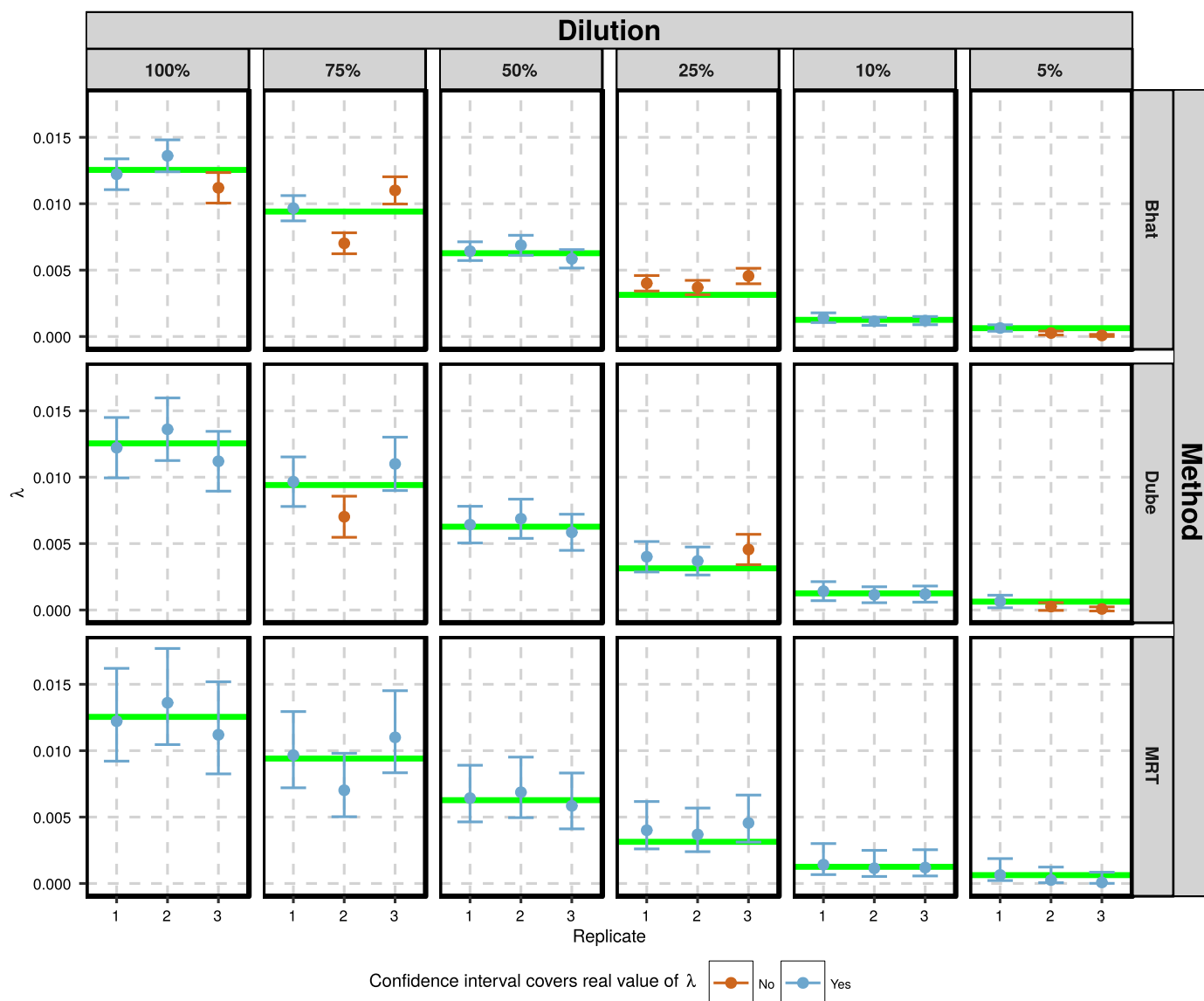


Figure 4. The confidence intervals of the results of the experiment. The red color marks confidence intervals that failed to cover the real value of λ denoted by a green line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the difference between a run with $\lambda = 0.01$ and runs with λ only up to 0.03 was not considered statistically significant by at least one method, whereas a run with $\lambda = 0.14$ differed significantly from runs with λ even up to 0.20.

The statistical power of the GLM and MRT methods depended also on the number of runs (Figs. 2 and S1, and Table S1). The GLM was more sensitive than MRT to changes of the template concentration when a single run per group was compared. In turn, the MRT had higher statistical power when tested scenarios involving more than one run per group. This is in line with the definition of a multiple comparison.

The speed of both approaches was assessed in a separate benchmark. The MRT method was always faster than GLM regardless of the number of runs compared. The difference considerably increased non-linearly with the number of runs, whereas the number of partitions had much smaller impact on the computation time (Fig. 3 and Table S2).

Coverage probability of confidence intervals

We performed 1×10^6 dPCR experiments *in silico* (2×10^4 droplets each) for the range of λ values from 0.05 to 2.00 with the increment 0.05 (1.2×10^7 experiments total).

To assess the simultaneous coverage probability, we randomly divided experiments into 2000 groups (500 experiments each) for all values of λ . We counted the frequency of groups in which all confidence intervals contained the true value of λ .

The adjusted confidence intervals, which are much wider than nominal/uncorrected CIs, guaranteed 0.95 stable simultaneous coverage probability. This offers more reliable comparison of multiple runs (Table S3).

Analysis of experimental data

We analyzed experimental data using the MRT. The proposed framework was able to distinguish between different concentrations even on restrictive confidence levels. As in the Monte Carlo evaluation, there are ‘blind spots’, where MRT is not able to discriminate between two relative concentrations, for example between 0% and 5% (Table 2, detailed results in Fig. S2 and Table S4).

MRT yields sufficiently high *p*-values when comparing the same concentrations with the only exception for the 75% concentration. In this case, only one sample was wrongly assessed as having the λ value different from other samples with the same concentration.

The confidence intervals determined by MRT always covered the real mean number of template molecules per partition. This indicates that all technical replicates come from the same source (Fig. 4 and Table S5). On the opposite, the real λ was not covered in

Table 2 Fraction of significant pairwise comparisons of dPCR results.

		Dilution (experiment 1)						
		0%	5%	10%	25%	50%	75%	100%
Dilution (experiment 2)	0%	0.00	0.33	1.00	1.00	1.00	1.00	1.00
	5%		0.00	0.67	1.00	1.00	1.00	1.00
	10%			0.00	1.00	1.00	1.00	1.00
	25%				0.00	0.67	1.00	1.00
	50%					0.00	0.67	1.00
	75%						0.67	0.44
	100%							0.00

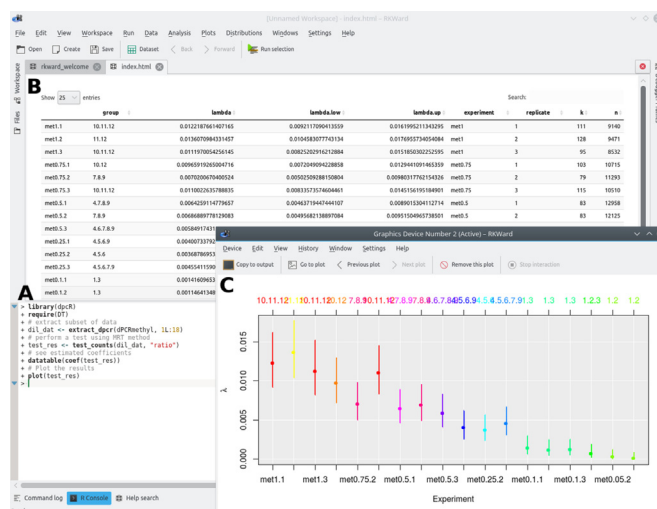


Figure 5. Example for the analysis of multiple dPCR runs. In this example, the graphical user interface Rkward [30] (v. 0.6.5) was used to compare multiple dPCR runs. The principle analysis of such experiments has been described in [15]. (A) Only a few lines of code are needed to load, process and report the results of the dPCR experiments. Here we used the dPCRmethy1 dataset contained in the dPCR package. (B) The results of the analysis can easily be presented as interactive table. (C) Plots of the analysis can be created as needed. In our example, we used the functionality of the dPCR package to create a plot automatically.

22% and 44% of cases by the confidence intervals computed using Dube’s and Bhat’s method, respectively.

Conclusions

The MRT framework seems to be a tailor-made method for comparing multiple dPCR experiments. This method is characterized by a statistical power which is superior to the simplistic GLM approach and is considerably faster. Performance is very important, because even the analysis of large experimental setups should be possible on average desktop computers.

Adjusted confidence intervals suitable for multiple comparison problems are an integral part of MRT. Although Dube’s method might be also enhanced by Dunn–Šidák correction for the purpose of multiple testing, it uses normal approximation to compute the confidence intervals. Others have shown that this approach is inferior to Wilson’s method as employed by MRT, because its probability coverage is inadequate for values of *k/n* close to 0 and 1 [24]. In consequence, for low and large values of λ , the real probability coverage of confidence intervals determined by Dube’s method may differ from its nominal value in contrast to the Wilson’s method.

It is worth to note that both analyzed statistical methods may be further enhanced. GLM may be expanded with more complex models taking more effects as recently shown into account [9]. MRT can utilize different tests for binomial proportion (e.g., test of equal proportions).

Both tests are available by the test_counts function of the dPCR package the commonly available open source statistical software environment R (see Fig. 5).

Acknowledgments

This work was supported by the InnoProfile-Transfer 03IPT611X (BMBF) and KMU-innovativ-16 031B0098B (BMBF) projects.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bdq.2016.06.004>.

References

- [1] R.A. White 3rd, P.C. Blainey, H.C. Fan, S.R. Quake, Digital PCR provides sensitive and absolute calibration for high throughput sequencing, *BMC Genomics* 10 (2009) 116, <http://dx.doi.org/10.1186/1471-2164-10-116>.
- [2] A.A. Morley, Digital PCR: a brief history, *Biomol. Detect. Quantif.* 1 (1) (2014) 1–2, <http://dx.doi.org/10.1016/j.bdq.2014.06.001> <http://www.sciencedirect.com/science/article/pii/S2214753514000023>.
- [3] J.M. Ruijter, M.W. Pfaffl, S. Zhao, A.N. Spiess, G. Boggy, J. Blom, R.G. Rutledge, D. Sisti, A. Lievens, K. De Preter, S. Derveaux, J. Hellemans, J. Vandesompele, Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications, *Methods (San Diego, CA)* 59 (1) (2013) 32–46, <http://dx.doi.org/10.1016/j.ymeth.2012.08.011>.
- [4] A.-N. Spiess, C. Deutschmann, M. Burdukiewicz, R. Himmelreich, K. Klat, P. Schierack, S. Rödiger, Impact of smoothing on parameter estimation in quantitative DNA amplification experiments, *Clin. Chem.* 61 (2) (2015) 379–388, <http://dx.doi.org/10.1373/clinchem.2014.230656> <http://www.clinchem.org/content/61/2/379>.
- [5] S. Rödiger, M. Burdukiewicz, P. Schierack, chipPCR: an R package to pre-process raw data of amplification curves, *Bioinformatics* 31 (17) (2015) 2900–2902, <http://dx.doi.org/10.1093/bioinformatics/btv205> <http://bioinformatics.oxfordjournals.org/content/31/17/2900.abstract>.
- [6] J.M. Ruijter, A. Ruiz Villalba, J. Hellemans, A. Untergasser, M.J.B. van den Hoff, Removal of between-run variation in a multi-plate qPCR experiment, *Biomol. Detect. Quantif.* 5 (2015), <http://dx.doi.org/10.1016/j.bdq.2015.07.001> <http://www.sciencedirect.com/science/article/pii/S2214753515300012>.
- [7] S. Pabinger, S. Rödiger, A. Kriegner, K. Vierlinger, A. Weinhäusel, A survey of tools for the analysis of quantitative PCR (qPCR) data, *Biomol. Detect. Quantif.* 1 (1) (2014) 23–33, <http://dx.doi.org/10.1016/j.bdq.2014.08.002> <http://www.sciencedirect.com/science/article/pii/S2214753514000059>.
- [8] J.F. Huggett, J. O'Grady, S. Bustin, qPCR, dPCR, NGS – a journey, *Biomol. Detect. Quantif.* 3 (2015) A1–A5, <http://dx.doi.org/10.1016/j.bdq.2015.01.001> <http://www.sciencedirect.com/science/article/pii/S2214753515000029>.
- [9] R.M. Dorazio, M.E. Hunter, Statistical models for the analysis and design of digital polymerase chain reaction (dPCR) experiments, *Anal. Chem.* 87 (21) (2015) 10886–10893, <http://dx.doi.org/10.1021/acs.analchem.5b02429>.
- [10] P.R. Debski, K. Gewartowski, M. Sulima, T.S. Kaminski, P. Garstecki, Rational design of digital assays, *Anal. Chem.* 87 (16) (2015) 8203–8209, <http://dx.doi.org/10.1021/acs.analchem.5b00942>.
- [11] N. Majumdar, T. Wessel, J. Marks, Digital PCR modeling for maximal sensitivity, dynamic range and measurement precision, *PLOS ONE* 10 (3) (2015) e0118833, <http://dx.doi.org/10.1371/journal.pone.0118833>.
- [12] S. Dube, J. Qin, R. Ramakrishnan, Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device, *PLoS ONE* 3 (8) (2008) e2876, <http://dx.doi.org/10.1371/journal.pone.0002876>.
- [13] S. Bhat, J. Herrmann, P. Armishaw, P. Corbisier, K.R. Emslie, Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number, *Anal. Bioanal. Chem.* 394 (2) (2009) 457–467, <http://dx.doi.org/10.1007/s00216-009-2729-5>.
- [14] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016 <https://www.R-project.org/>.
- [15] S. Rödiger, M. Burdukiewicz, K.A. Blagodatikh, P. Schierack, R as an environment for the reproducible analysis of DNA amplification experiments, *R. J.* 7 (2) (2015) 127–150 <http://journal.r-project.org/archive/2015-1/RJ-2015-1.pdf>.
- [16] J.J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, CRC Press, 2016.
- [17] A. Zeileis, C. Kleiber, S. Jackman, Regression models for count data in R, *J. Stat. Softw.* 27 (8) (2008) 1–25 <http://www.jstatsoft.org/v27/i08>.
- [18] G. Hommel, F. Bretz, W. Maurer, Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies, *Stat. Med.* 26 (22) (2007) 4063–4073, <http://dx.doi.org/10.1002/sim.2873>.
- [19] F. Bretz, T. Hothorn, P. Westfall, *Multiple Comparisons Using R*, Chapman & Hall/CRC Press, Boca Raton, FL, USA, 2010.
- [20] K.F. Hirji, *Exact Analysis of Discrete Data*, Chapman & Hall/CRC, 2005.
- [21] M. Fay, Two-sided exact tests and matching confidence intervals for discrete data, *R. J.* 2 (1) (2010) 53–58.
- [22] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B (Methodol.)* 57 (1) (1995) 289–300, <http://dx.doi.org/10.2307/2346101>.
- [23] B.E. Storer, C. Kim, Exact properties of some exact test statistics for comparing two binomial proportions, *J. Am. Stat. Assoc.* 85 (409) (1990) 146–155, <http://dx.doi.org/10.1080/01621459.1990.10475318>.
- [24] L.D. Brown, T.T. Cai, A. DasGupta, Interval estimation for a binomial proportion, *Stat. Sci.* 16 (2) (2001) 101–133, <http://dx.doi.org/10.1214/ss/1009213286>.
- [25] Z. Šidák, Rectangular confidence regions for the means of multivariate normal distributions, *J. Am. Stat. Assoc.* 62 (318) (1967) 626–633, <http://dx.doi.org/10.1080/01621459.1967.10482935>.
- [26] M. Burdukiewicz, S. Rödiger, B. Jacobs, P. Sobczyk, dpcR: Digital PCR Analysis, R Package Version 0.2, 2015 <http://CRAN.R-project.org/package=dpcR>.
- [27] T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models, *Biom. J.* 50 (3) (2008) 346–363.
- [28] M. Fay, rateratio.test: Exact Rate Ratio Test, R Package Version 1.0-2, 2014 <https://CRAN.R-project.org/package=rateratio.test>.
- [29] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer, New York, 2009 <http://had.co.nz/ggplot2/book>.
- [30] S. Rödiger, T. Friedrichsmeier, P. Kapat, M. Michalke, Rkward: a comprehensive graphical user interface and integrated development environment for statistical analysis with R, *J. Stat. Softw.* 49 (9) (2012) 1–34 <http://www.jstatsoft.org/v49/i09>.